DOCUMENT RESUME

ED 420 694 TM 028 369

AUTHOR Roberts, James S.; Wedell, Douglas H.; Laughlin, James E.
TITLE Heightened Sensitivity of Likert Attitude Scales to Restriction of Sample Range.
PUB DATE 1998-04-17
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Attitudes; *Likert Scales; *Sampling; Test Items; *Validity
IDENTIFIERS *Item Characteristic Function; Thurstone Scales

ABSTRACT

The Likert rating scale procedure is often used in conjunction with a graded disagree-agree response scale to measure attitudes. Item characteristic curves associated with graded disagree-agree responses are generally single-peaked, nonmonotonic functions of true attitude. These characteristics are, thus, more generally consistent with an unfolding model like the Thurstone attitude measurement procedure rather than a cumulative model such as the Likert procedure and the typical item characteristics of the disagree-agree responses can lead to invalid measurement such that individuals with the most extreme attitudes receive Likert scores that are indicative of more moderate opinions. This type of invalidity may be exacerbated when a Likert scale is constructed under restrictions of the sample attitude range. Specifically, a restricted sample range may mask the nonmonotonic response characteristics of a given item so that its characteristic curve appears to be monotonically related to attitude. The item would, thus, appear worthy of inclusion on a Likert scale. If that Likert scale was subsequently used with a broader item sample, the nonmonotonic characteristics of the item would reappear, and the validity problems inherent in measuring individuals with the most extreme attitudes would intensify. (Contains 4 figures, 2 tables and 23 references.) (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 420 694

Heightened Sensitivity of Likert Attitude Scales to Restriction of Sample Range

James S. Roberts
Medical University of South Carolina

Douglas H. Wedell and James E. Laughlin
University of South Carolina

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

James Roberts

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Presented at the 1998 American Educational Research Association Annual Meeting
April 17, San Diego, California

TM028369

To find out more about new unfolding models for attitude measurement, visit our Web site at www.musc.edu/cdap/roberts

Abstract

The Likert rating scale procedure is often used in conjunction with a graded disagree-agree response scale to measure attitudes. Item characteristic curves associated with graded disagree-agree responses are generally single-peaked, nonmonotonic functions of true attitude. These characteristics are, thus, more generally consistent with an unfolding model like the Thurstone attitude measurement procedure rather than a cumulative model such as the Likert procedure. Previous research has illustrated that the inconsistency between the Likert procedure and the typical item characteristics of disagree-agree responses can lead to invalid measurement such that individuals with the most extreme attitudes receive Likert scores that are indicative of more moderate opinions. This type of invalidity may be exacerbated when a Likert scale is constructed under restrictions of the sample attitude range. Specifically, a restricted sample range may mask the nonmonotonic response characteristics of a given item so that its characteristic curve appears to be monotonically related to attitude. The item would, thus, appear worthy of inclusion on a Likert scale. If that Likert scale was subsequently used with a broader attitude sample, the nonmonotonic characteristics of the item would reappear, and the validity problems inherent in measuring individuals with the most extreme attitudes would intensify.

Several researchers have recently argued that binary or graded disagree-agree responses to attitude questionnaire items are generally more consistent with an unfolding model as opposed to a cumulative model (Andrich, 1996; Roberts, 1995; Roberts, Laughlin & Wedell, 1998; van Schuur & Kiers, 1994). An unfolding model suggests that an individual will endorse an attitude statement to the extent that the statement reflects the individual's own opinion about the issue in question. Specifically, an unfolding model represents a proximity relation (Coombs, 1964) in which the individual endorses an item to the extent that the individual and the item are located relatively close to each other on a latent attitude continuum. Figure 1 illustrates three theoretical item characteristic curves that one would expect for a moderately negative, a neutral and a moderately positive attitude statement under a unidimensional unfolding model. Each characteristic curve is nonmonotonic and reaches a maximal level of expected agreement when a given individual's attitude level matches the attitude expressed by the statement (i.e., when the individual and the item are located close to each other on the underlying attitude continuum). The Thurstone (1928) attitude measurement procedure is consistent with item characteristics like those portrayed in Figure 1 (Green, 1954), and in that sense, it is implicitly a type of unfolding model. Furthermore, there have been several parametric (Andrich, 1988, 1996; Andrich & Luo, 1993; Hoijtink, 1990,1991; Roberts, 1995; Roberts & Laughlin, 1996ab; Roberts, Donoghue & Laughlin, 1996) and nonparametric (Cliff, Collins, Zatkin, Gallipeau & McCormick, 1989; van Schuur, 1984, 1993) item response models proposed for measuring attitudes using an unfolding model for disagree-agree responses.

Insert Figure 1 About Here

The Likert (1932) rating scale procedure is often used in conjunction with a graded disagree-agree response scale to measure attitudes. However, a Likert scale is implicitly a type of cumulative model (Green, 1954; Roberts, 1995; Roberts et al., 1998). A cumulative model represents a dominance relation in which the individual endorses an item to the extent that the individual's location on the attitude continuum exceeds that for the item (Coombs, 1964). Figure 2 illustrates theoretical item characteristic curves for a moderately negative, a neutral and a moderately positive attitude item under a unidimensional cumulative model. Note that after reverse scoring responses to negatively worded items, the cumulative model suggests greater expected levels of agreement to the extent that an individual's attitude dominates the sentiment expressed by the item (i.e., to the degree that the individual's location on the attitude continuum exceeds the location of the item).

Insert Figure 2 About Here

If disagree-agree responses to attitude statements generally follow from a proximity relation, and are, thus, consistent with an unfolding model, then why does an implicitly cumulative model

like the Likert scaling procedure work as well as it typically does? The answer to this question lies in how items for a Likert scale are chosen. The typical item selection procedures used in the Likert scale construction process (e.g., item-total correlations, discrimination indices, loadings from the first principal component, indices based on Cronbach's alpha) favor those items with monotonic item characteristic curves. Moreover, even when responses follow from a proximity relation, those items which express extreme opinions relative to the individual attitudes in the sample will exhibit item characteristic curves that more or less approximate a monotonic function. For example, the trace lines shown for the moderately positive and moderately negative items in Figure 1 are monotonic if one ignores the extreme regions of the attitude continuum. The Likert scale construction process tends to eliminate the neutral attitude items in favor of those that are more extreme and exhibit roughly monotonic trace lines (Roberts, 1995; Roberts et al., 1998). This accounts for a historically consistent finding in the attitude measurement literature: Likert scales are typically void of neutral items (Andrich, 1996; Edwards, 1946; Edwards and Kenney, 1946; Ferguson, 1941).

Roberts et al. (1998) have shown that it can be difficult to obtain items with completely monotonic characteristic curves when a Likert scale is constructed from disagree-agree responses to attitude items. Even when using relatively extreme statements, the item characteristic curves might exhibit some nonmonotonic behavior in the extreme attitude regions as illustrated in Figure 1. Theoretically, this is because a statement may not be extreme enough to attract a consistently high degree of endorsement from individuals whose attitudes are even more extreme than the sentiment expressed by the statement. Roberts et al. have argued that when items on a Likert scale exhibit nonmonotonic behavior in the extreme regions of the attitude continuum, then the validity of attitude measurements can be degraded. Specifically, individuals with the most extreme true attitudes can receive scores that are indicative of more moderate opinions.

This paper will illustrate how the item characteristic curves developed from disagree-agree responses are highly dependent on the range of attitudes in a given sample. In particular, items with truly nonmonotonic characteristic curves may appear to have monotonic characteristics when the sample range of individual attitudes is restricted, and thus, such items would more likely be selected for a Likert attitude scale. If the same Likert scale was subsequently used to measure attitudes in a broader sample of opinions, then the measurement problems identified by Roberts et al. (1998) would be magnified - those individuals with the most extreme attitudes would receive scores that seriously underestimated the extremity of their opinions. Consequently, the validity of Likert attitude measures can be quite sensitive to the range of sample attitudes measured in a given application relative to the range of attitudes studied during the scale construction process.

Method

Development of Item Characteristic Curves

Empirical item characteristic curves were developed for 10 statements that reflected a wide

range of attitudes toward abortion. The curves were based on data previously reported by Roberts et al. (1998). Subjects were 301 University of South Carolina undergraduates who rated each statement with regard to its unfavorability/favorability toward abortion using a 9-point scale. These favorability ratings were used as input to the successive intervals procedure (Safir, 1937), from which Thurstone scale values and associated standard deviations were derived for each statement. Additionally, graded disagree-agree responses from 781 subjects were obtained for each item using a 6-point scale where 1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree, 4=Slightly Agree, 5=Agree, and 6= Strongly Agree. Thurstone attitude scores were developed for each of these subjects by computing the median scale value associated with those statements that a given individual endorsed to at least some extent (i.e., those statements endorsed with a response of 4 or more). Subjects were arranged in order of their Thurstone attitude score and then divided into 26 relatively homogeneous attitude score groups with approximately 30 persons per group. Empirical item characteristic curves were calculated for each item by plotting the mean item response against the mean attitude score within each group.

Development of Likert Attitude Scales Under Censored and Uncensored Sample Conditions

The graded disagree-agree responses from the 781 subjects were used to develop three different Likert scales from the 10 items in the initial pool. Items for each scale were chosen on the basis of three criteria. First, each scale item had to have an absolute pattern value of .4 or greater on the first principal component of the interitem correlation matrix. Second, each scale item had to have an item-total correlation (after reverse scoring negatively worded items) of .35 or greater. Third, removal of a given scale item could not increase the Cronbach's alpha computed for the remaining items. Items that failed to meet these criteria were removed one-by-one in an iterative fashion such that the most internally inconsistent item was removed at a given step.

Selection of items for the first Likert scale was based on the responses of all 781 subjects (the uncensored condition). A second Likert scale was developed by repeating the item selection process after censoring those cases from the lowest 25% of the Thurstone attitude score distribution (the left-censored condition). Finally, items for the third Likert scale were selected after censoring the highest 25% of the Thurstone attitude score distribution (the right-censored condition). Scores from the three Likert scales were compared to each other when all sample restrictions were removed. They were also compared to the original Thurstone attitude scores.

Results

Form of Item Characteristic Curves

The item characteristic curves associated with the 10 attitude statements are illustrated in Figure 3. The curves are ordered according to their sentiment (and corresponding Thurstone scale value) from very negative to very positive content with regard to advocacy for the practice

of abortion. The most obvious feature of these curves is that the graded disagree-agree responses to each item are consistent with a proximity relation rather than a dominance relation. Individuals with the most negative Thurstone attitude scores endorsed the extremely negative statements (Items 1 and 2) the most, while those persons with neutral and positive attitudes toward abortion endorsed these items successively less. The reverse finding emerged for the extremely positive attitude statements (Items 9 and 10), which were endorsed most by individuals with the most positive attitudes, and endorsed successively less as individual attitudes became moderate and then negative, respectively. Most importantly, the moderately negative statements (Items 3 and 4), the neutral statements (Items 5 and 6) and the moderately positive statements (Items 7 and 8) all showed "folded" characteristic curves. Specifically, individuals with attitudes that were well-reflected by these statements endorsed them the most, but other persons with more extreme attitude positions, whether in a negative or positive direction, endorsed these items to a much lesser extent. Thus, these items exhibited characteristic curves that were markedly nonmonotonic. All curves were more or less single-peaked and consistent with the notion of a proximity relation. This pattern would suggest that an unfolding model, rather than a cumulative model, would be most appropriate for these data.

Insert Figure 3 About Here

It should be noted that the form of the characteristic curves is not an artifact of the Thurstone scoring procedure. It is true that in the Thurstone scale construction process, items with characteristic curves like those shown in Figure 3 are deemed as "relevant" to the attitude under study (Thurstone, 1928; Thurstone and Chave, 1929), and well-suited for an attitude scale. Therefore, a researcher designing a Thurstone scale would deliberately search for items with characteristics like those in Figure 3. However, there is nothing inherent in the Thurstone scoring procedure that artifactually produces items with characteristic curves like those in Figure 3. The Thurstone scoring procedure simply takes the median or mean of the scale values associated with those items an individual has endorsed. Assuming that the Thurstone scale values are good estimates of the true scale values, then the ultimate Thurstone attitude scores should be monotonically related to true attitudes regardless of whether the data are consistent with a proximity or a dominance relation. For example, if a dominance relation holds, then a given individual should, in all probability, cumulatively endorse successively more positive items as the individual's attitude becomes increasingly more positive. If one then takes the median of the item scale values that the individual endorses, this value should logically grow larger as successively more positive statements are endorsed. Theoretically, the resulting scores should be monotonically related to the individual's true attitude. In contrast, if a proximity relation holds, then a given individual should, in all probability, reduce endorsement for neutral and, perhaps, slightly favorable items as the individual's attitude becomes increasingly more positive. However, the scale values of the endorsed items will still indicate the position of the individual on the attitude continuum, and thus, the median of these scale values should still be monotonically related to the individual's attitude. Theoretically, the median scale value of endorsed items should

provide an ordinal estimate of true attitude in either situation, and therefore, the Thurstone score should enable a researcher to estimate an item characteristic curve regardless of whether the item responses are consistent with a cumulative or an unfolding model.

Form of Item Characteristic Curves Under Uncensored and Censored Sample Conditions

The vertical lines shown on each plot in Figure 3 indicate the 25th and 75th percentiles of the Thurstone attitude score distribution. The form of the characteristic curves associated with the moderate and neutral items was clearly nonmonotonic. However, if one censored the attitude distribution at either the 25th percentile (the left-censored condition) or the 75th percentile (the right-censored condition), then the form of these characteristic curves changed dramatically. For example, responses to Items 3 through 6 approximated a monotonically decreasing function of Thurstone attitude scores under the left-censoring condition. (To see this, simply cover-up that portion of each characteristic curve to the left of the 25th percentile line.) The same finding emerged to a slight extent for Items 7 and 8, however, these two items exhibited relatively less response variation as compared to Items 3 through 6. Under the right censoring condition, in contrast, responses to Items 5 through 8 approximated a monotonically increasing function of Thurstone attitude scores while responses to Items 3 and 4 appeared nonmonotonically related to attitudes.

Naturally, the extreme attitude items with the originally monotonic characteristic curves (i.e., Items 1, 2, 9 and 10) were still monotonic in appearance after either right-censoring or left-censoring. The censoring, thus, had little impact on the basic form of the item characteristics for extreme items, but had tremendous impact on the interpretation of item characteristics for moderate and neutral items. In other words, censoring had little influence on the basic form of curves that were originally monotonic, but it had substantial influence on how originally nonmonotonic characteristic curves would ultimately be interpreted.

Selection of Likert Scale Items Under Uncensored and Censored Sample Conditions

Likert item selection procedures identify those items that appear to be monotonically, if not linearly, related to the underlying trait. Given the changes in item characteristic curves produced by alternative censoring conditions, it was not surprising to find that the Likert scales developed in the alternative conditions varied with regard to the neutral and moderate items included on a given scale. The Likert scale developed in each sample censoring condition contained 8 of the original 10 items. In the uncensored condition, the two neutral items (Items 5 and 6) with the most strikingly nonmonotonic characteristics were eliminated, and the remaining 8 items were used to form the Likert scale. In the left-censored condition, the two moderately positive items (Items 7 and 8) were discarded from the Likert scale. These items still appeared more or less monotonic under left-censoring conditions, but they exhibited less internal consistency relative to the items that remained on the Likert scale. Finally, the two moderately negative items were eliminated from the Likert scale in the right-censored condition because these items maintained markedly nonmonotonic characteristic curves under this condition.

Insert Table 1 About Here

Table 1 shows the item selection indices associated with the 8 items maintained on each Likert scale. The indices were calculated within the sample condition in which a given scale was constructed (i.e., using either an uncensored, a left-censored or a right censored sample). The indices suggested that each scale was relatively homogeneous and internally consistent. For example, the Cronbach's alpha for the uncensored, left-censored and right-censored scales was equal to .87, .81 and .88, respectively. Additionally, the proportion of item response variance explained by the first principal component was equal to .53, .43, and .55 for the same three scales. If this were an actual scale construction situation, then the indices reported in Table 1 would obviously need to be cross-validated on an independent sample. However, the indices suggest that any of these scales would warrant consideration as a potential attitude measurement instrument.

Characteristics of Likert Scores When the Sample is Broadened

Scores from the Likert scales developed under the 3 sample censoring conditions were compared when all censoring restrictions were removed. Pearson and Spearman correlations between these scores are given in Table 2. Scores from the Likert scales constructed under the right-censored and left-censored conditions were highly correlated with those from the Likert scale developed in the uncensored condition ($r > .85$). However, the scores from the right-censored and left-censored scales were, at best, moderately correlated with each other ($r < .5$).

Insert Table 2 About Here

Scores from the three Likert scales are plotted against the original Thurstone attitude scores in the top panel of Figure 4. Scores from the Likert scale developed in the uncensored sample condition exhibited the nonmonotonic relationship with Thurstone attitude scores previously illustrated by Roberts et al. (1998). Specifically, those individuals with the most extreme Thurstone attitude scores (either extremely negative or extremely positive) received Likert scores that were indicative of slightly more moderate attitudes. This result was studied further by independently ranking both the Likert and Thurstone scores, and then comparing the resulting ranks for the most extreme Thurstone attitude groups using a Wilcoxon signed rank test. For the most negative Thurstone attitude group, the median rank for Thurstone scores was equal to 24, whereas that for Likert scores was 65. This difference was statistically significant ($S = +226.5$, $p < .0001$). For the most positive Thurstone attitude group, the median rank for Thurstone scores was equal to 764 and that for Likert scores was 532.5. Again, this difference was statistically significant ($S = -245.0$, $p < .0001$). Thus, the uncensored Likert scale suggested that individuals

with the most extreme Thurstone attitude scores had relatively more moderate opinions, and although this effect was slight, it was statistically significant.

The Likert scales developed under censoring conditions exacerbated the findings observed with the uncensored scale in a predictable fashion. The Likert scale derived in the left-censored condition magnified the nonmonotonic relationship between the Likert and Thurstone scores in the negative (left) region of the attitude continuum when the sample was broadened. The Likert scale developed in the right-censored condition intensified the nonmonotonic relationship between Likert and Thurstone scores in the positive (right) region of the attitude continuum once the sample restrictions were removed. These findings generally became even more apparent when ranks of Likert scores were plotted against Thurstone scores as shown in the bottom panel of Figure 4.

Insert Figure 4 About Here

Discussion

Roberts et al. (1998) have previously suggested that cumulative models such as the Likert procedure may substantially underestimate the extremity of the attitudes for individuals with the most extreme opinions when estimates are based on binary or graded disagree-agree responses. This underestimation can lead to a reversal of the ordering of individuals on the latent trait so that people with the most extreme attitudes receive more moderate scores than those with less extreme attitudes. The results from these analyses imply that the magnitude of underestimation can be exacerbated when a Likert scale is constructed under constrained sampling conditions and then applied to a broader sample. Specifically, when a sample from an attitude distribution is constrained, an item with basically nonmonotonic characteristics may appear to be monotonically related to the latent attitude. If that item is then used to measure attitudes with a broader sample, its nonmonotonic characteristics will emerge. It is the nonmonotonicity in item characteristic curves that, in turn, leads to the underestimation of extreme attitudes illustrated by Roberts et al. The degree to which this occurs in the real world is unknown. However, given that Likert attitude questionnaires are often developed from relatively homogeneous samples (e.g., college students, treatment seekers, special interest groups, etc.), it is certainly possible that the problems illustrated in this report may surface in real measurement applications.

It is interesting to contrast the effects that restriction of sample range may have when responses follow an unfolding model as opposed to a cumulative model. Under a cumulative model, item-trait correlations generally attenuate when the sample range is restricted due to the fact that an essentially monotonic item characteristic curve is truncated. Consequently, there is less variability in the observed item responses that can be linearly related to the latent trait. Therefore, item-trait correlations typically decrease when samples are censored. In contrast, item-

trait correlations may decrease, increase or even completely reverse when responses follow an unfolding model and the sample range is restricted. This is due to the fact that item characteristic curves may be chiefly monotonic or markedly nonmonotonic depending on the location of the item relative to the individuals in the sample. If the item characteristic curve is more or less monotonic throughout the attitude continuum, then restriction of sample range will generally lead to decreased (attenuated) item-trait correlations as in the cumulative model. In contrast, if the item characteristic curve is truly nonmonotonic across the attitude continuum, then restriction of sample range may increase the apparent monotonicity of the characteristic function, and consequently, increase the absolute item-trait correlation. Furthermore, whether a truly nonmonotonic item characteristic curve appears, after censoring, to be basically an increasing or decreasing function of attitude depends on the region of the attitude distribution that is restricted (i.e., it depends on whether negative attitudes or positive attitudes are censored). Thus, the item-trait correlation can vary from positive to negative depending on the type of sample restrictions that occur. In summary, the consequences of sample restrictions on item-trait correlations appear to be far more complex when data follow from a proximity relation.

This report has further illustrated some of the undesirable problems that can emerge when using a cumulative model to estimate attitudes from binary or graded disagree-agree responses. The generality and typical magnitude of these problems remain to be studied, but attitude researchers should, nonetheless, be sensitive to the basic measurement issues from which the problems emerge. We recommend that attitude researchers consider using unfolding models to develop attitude estimates from binary or graded disagree-agree responses. These procedures generally allow the researcher to examine item characteristic curves and thus gain greater insight into how such responses should be interpreted. The classical Thurstone model is an example of an unfolding model, but there are more contemporary unfolding item response models to choose from. Our research has clearly demonstrated that development of a Likert attitude scale on a sample with a restricted range can lead to a scale that is invalid when applied to a broader population. Therefore, at the very least, attitude researchers should attempt to secure the broadest possible attitude sample when developing a Likert attitude scale in order to avoid nonmonotonic items that will lead to invalid conclusions.

References

Andrich, D. (1988). The application of an unfolding model of the PIRT type to the measurement of attitude. Applied Psychological Measurement, 12, 33-51.

Andrich, D. (1996). A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. British Journal of Mathematical and Statistical Psychology, 49, 347-365.

Andrich, D., & Luo, G. (1993). A hyperbolic cosine latent trait model for unfolding dichotomous single-stimulus responses. Applied Psychological Measurement, 17, 253-276.

Cliff, N., Collins, L. M., Zatkin, J., Gallipeau, D., & McCormick, D. J. (1988). An ordinal scaling method for questionnaire and other ordinal data. Applied Psychological Measurement, 12, 83-97.

Coombs, C. H. (1964). A theory of data. New York: Wiley.

Edwards, A. L. (1946). A critique of 'neutral' items in attitude scales constructed by the method of equal appearing intervals. Psychological Review, 53, 159-169.

Edwards, A. L., & Kenney, K. C. (1946). A comparison of the Thurstone and Likert techniques of attitude scale construction. Journal of Applied Psychology, 30, 72-83.

Ferguson, L. W. (1941). A study of the Likert technique of attitude scale construction. Journal of Social Psychology, 13, 51-57.

Green, B. F. (1954). Attitude measurement. In G. Lindzey (Ed.), Handbook of social psychology (1st ed., Vol. 1, pp. 335-369). Cambridge, MA: Addison-Wesley.

Hoijtink, H. (1990). A latent trait model for dichotomous choice data. Psychometrika, 55, 641-656.

Hoijtink (1991). The measurement of latent traits by proximity items. Applied Psychological Measurement, 15, 153-169.

Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, No. 140, 5-53.

Roberts, J. S. (1995). Item Response Theory Approaches to Attitude Measurement. (Doctoral dissertation, University of South Carolina, Columbia, 1995). Dissertation Abstracts International, 56, 7089B.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (1996). A generalized item response model for unfolding responses from a graded scale. Paper presented at the 61st annual meeting of the Psychometric Society, June 28, Banff, Alberta, Canada.

Roberts, J. S., & Laughlin, J. E. (1996a). A unidimensional item response model for unfolding responses from a graded disagree-agree response scale. Applied Psychological Measurement, 20, 231-255.

Roberts, J. S., & Laughlin, J. E. (1996b). The graded unfolding model: A unidimensional item response model for unfolding graded responses (RR-96-16). Princeton, NJ: Educational Testing Service.

Roberts, J. S, Laughlin, J. E., & Wedell, D. H. (1998). Validity issues in the Likert and Thurstone Approaches to Attitude Measurement. Educational and Psychological Measurement, in press.

Safir, M. A. (1937). A comparative study of scales constructed by three psychophysical methods. Psychometrika, 2, 179-198.

Thurstone, L. L. (1928). Attitudes can be measured. The American Journal of Sociology, 26, 249-269.

Thurstone, L. L., & Chave, E. J. (1929). The measurement of attitude: A psychophysical method and some experiments with a scale for measuring attitude toward the church. Chicago: University of Chicago Press.

van Schuur, W. H. (1984). Structure in political beliefs: A new model for stochastic unfolding with application to European party activists. Amsterdam: CT Press.

van Schuur, W. H. (1993). Nonparametric unidimensional unfolding for multicategory data. Political Analysis, 4, 41-74.

van Schuur, W. H., & Kiers, H. A. L. (1994). Why factor analysis is often the incorrect model for analyzing bipolar concepts, and what model can be used instead. Applied Psychological Measurement, 18, 97-110.

Table 1. Item selection indices for Likert scales constructed under uncensored, left-censored and right-censored conditions

	Absolute Pattern Value from First Principal Component			Corrected Item-Total Correlation			Cronbach's Alpha if Item Deleted		
Item	UC	LC	RC	UC	LC	RC	UC	LC	RC
1.	.82	.53	.83	.74	.39	.75	.84	.80	.86
2.	.77	.71	.60	.68	.56	.50	.85	.78	.88
3.	.64	.80	NA	.53	.67	NA	.86	.76	NA
4.	.65	.79	NA	.55	.66	NA	.86	.76	NA
5.	NA	.73	.63	NA	.61	.53	NA	.77	.88
6.	NA	.66	.63	NA	.54	.53	NA	.78	.88
7.	.58	NA	.84	.47	NA	.77	.87	NA	.86
8.	.67	NA	.83	.56	NA	.76	.86	NA	.86
9.	.81	.45	.76	.72	.35	.67	.84	.81	.87
10.	.83	.49	.78	.75	.38	.70	.84	.80	.86

Item numbers correspond to those used in Figure 3. UC=uncensored condition, LC=left-censored condition, RC=right-censored condition, NA=not applicable because given item was not on scale.

Table 2. Pearson and Spearman correlations between scores from Likert scales constructed under uncensored, left-censored and right-censored conditions after removing all sample restrictions.

	Pearson Correlation (r)			Spearman Correlation (r_s)		
	UC	LC	RC	UC	LC	RC
UC	1.0	.86	.86	1.0	.90	.75
LC	-	1.0	.49	-	1.0	.46
RC	-	-	1.0	-	-	1.0

UC=uncensored condition, LC=left-censored condition, RC=right-censored condition

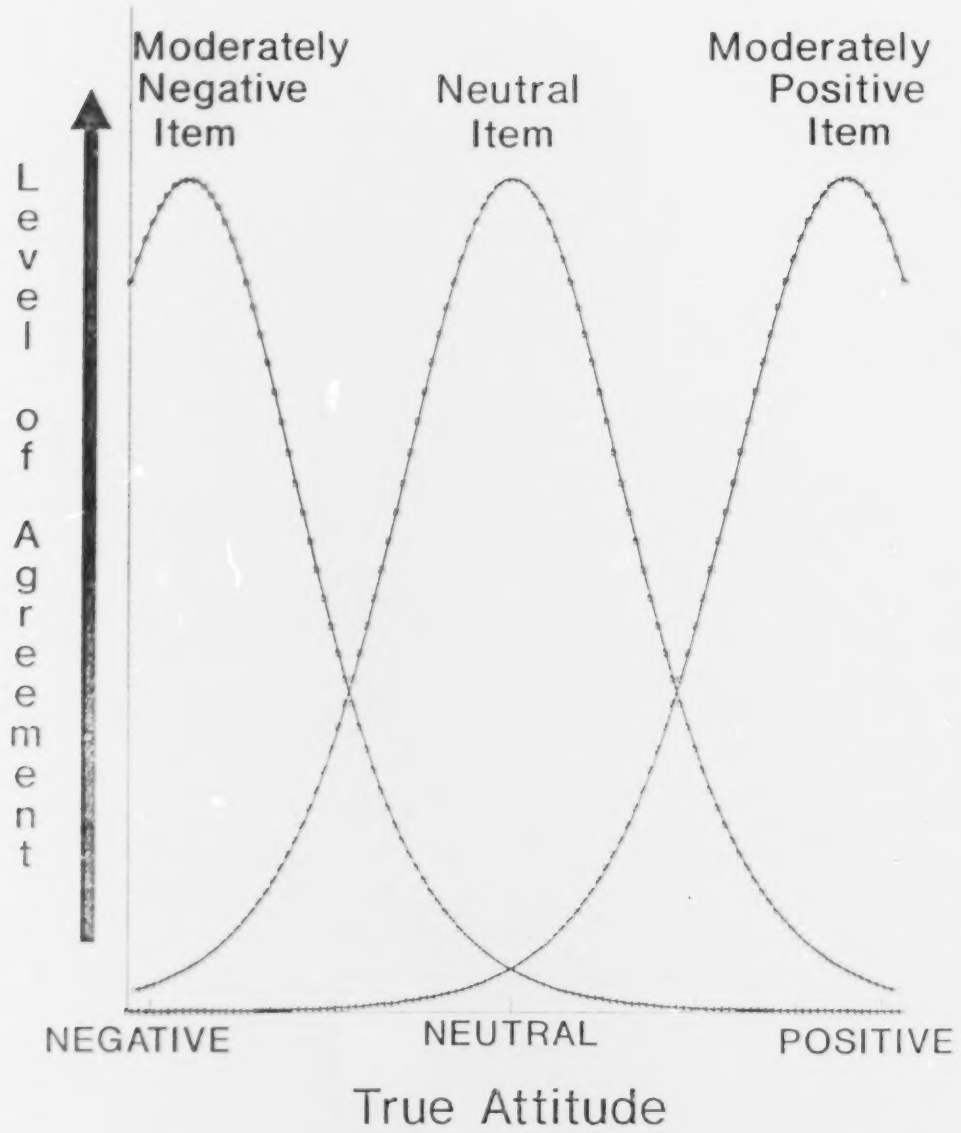


Figure 1. Theoretical item characteristic curves corresponding to an unfolding item response model. The theoretical level of agreement is shown for a moderately negative, a neutral and a moderately positive attitude statement. The unfolding item response model suggests that an individual will agree with an item to the extent that the individual is located near the item on the latent attitude continuum.

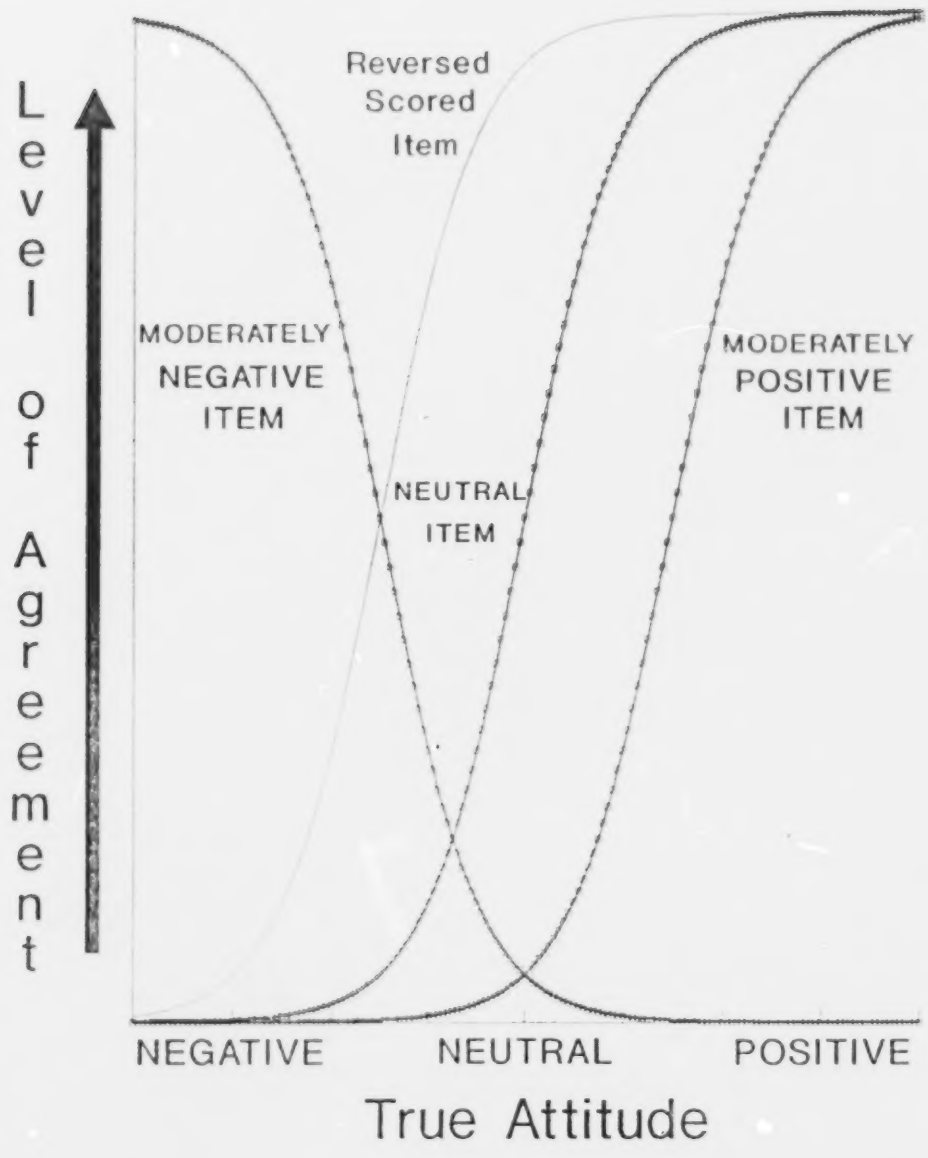


Figure 2. Theoretical item characteristic curves corresponding to a cumulative item response model. The theoretical level of agreement is shown for a moderately negative, a neutral and a moderately positive attitude statement. After reverse scoring negatively worded items, the cumulative item response model suggests that an individual will agree with an item to the extent that the individual's location exceeds the item's location on the latent attitude continuum.

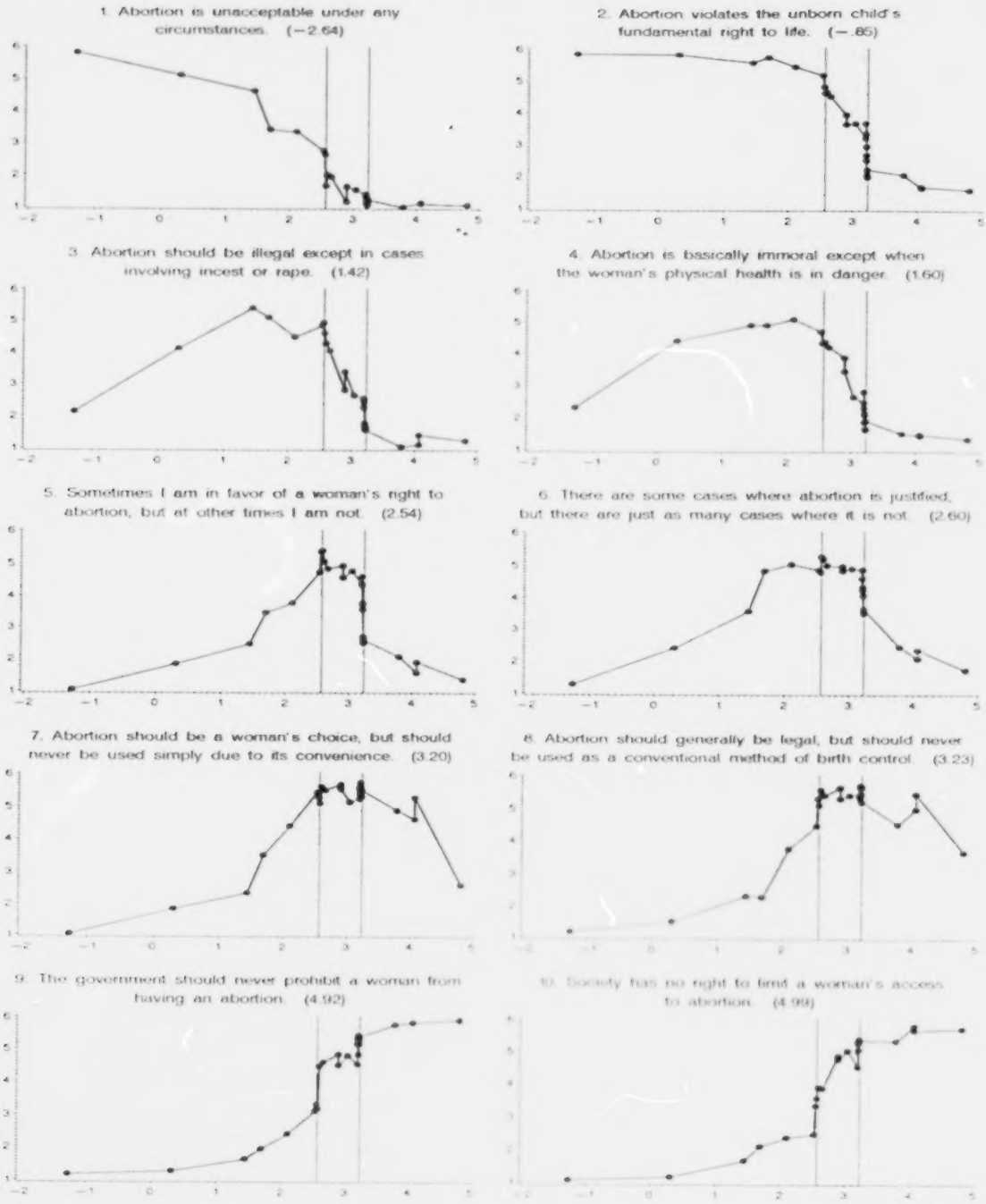
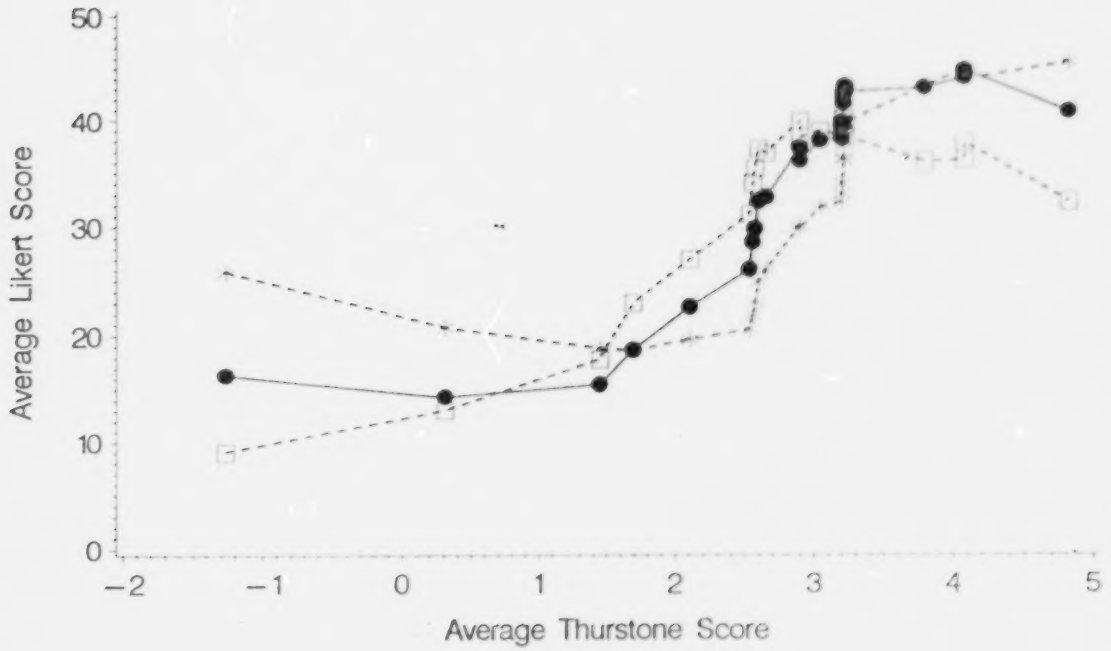

1. Abortion is unacceptable under any circumstances. (−2.64)
2. Abortion violates the unborn child's fundamental right to life. (−.85)
3. Abortion should be illegal except in cases involving incest or rape. (1.42)
4. Abortion is basically immoral except when the woman's physical health is in danger. (1.60)
5. Sometimes I am in favor of a woman's right to abortion, but at other times I am not. (2.54)
6. There are some cases where abortion is justified, but there are just as many cases where it is not. (2.60)
7. Abortion should be a woman's choice, but should never be used simply due to its convenience. (3.20)
8. Abortion should generally be legal, but should never be used as a conventional method of birth control. (3.23)
9. The government should never prohibit a woman from having an abortion. (4.92)
10. Society has no right to limit a woman's access to abortion. (4.99)

Figure 3. Empirical item characteristic curves associated with 10 items designed to measure attitudes toward abortion. Estimated Thurstone scale values are given parenthetically. Each vertical axis denotes the mean level of observed agreement (1="Strongly Disagree", ..., 6="Strongly Agree"), whereas each horizontal axis denotes the mean Thurstone attitude score. Means were calculated within homogeneous Thurstone attitude score groups composed of approximately 30 respondents per group. The vertical lines indicate the 25th and 75th percentiles of the Thurstone score distribution.



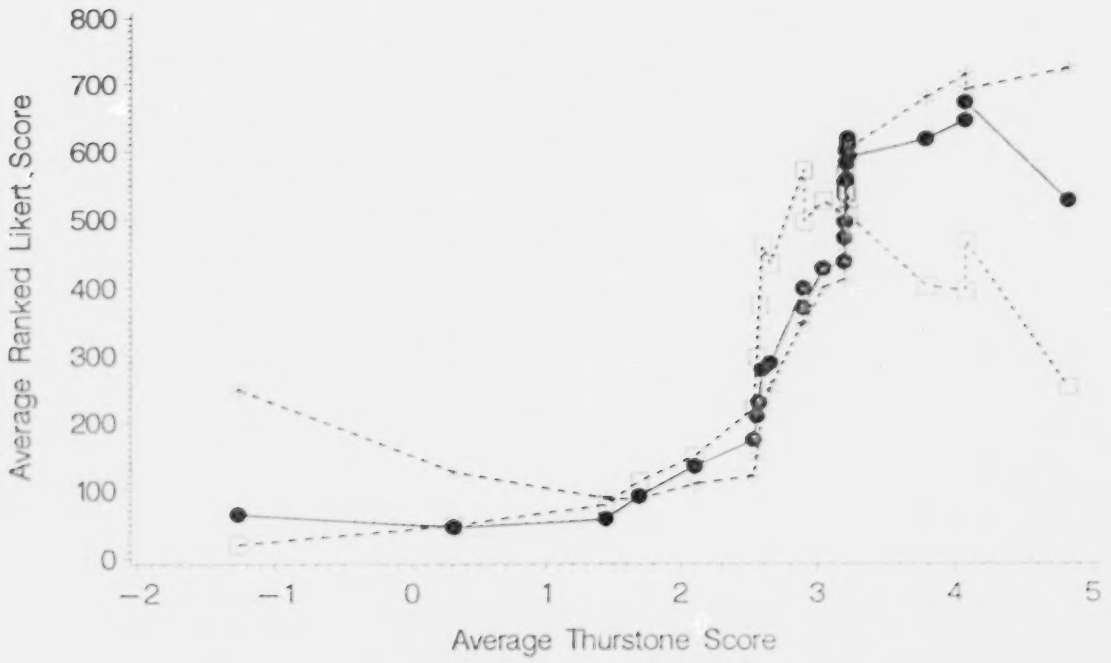


Figure 4. Relationships between average Thurstone scores, average Likert scores (top panel), and average ranked Likert scores (bottom panel). Averages were calculated within homogeneous Thurstone attitude score groups composed of approximately 30 respondents per group. Likert scores were based on scales originally constructed under right-censored, left-censored and uncensored sample conditions (denoted by □, *, and ●, respectively).